

口语分班测试题型研究*

李海燕 蔡云凌 刘颂浩

提要 本文在以往口语分班测试研究的基础上,对六种题型(自我介绍、对话、话题表达、看图说话、朗读以及听读)进行了研究,目的是回答两个问题:1)口语分班测试采用哪些题型或题型组合更有效?2)对不同汉字背景的学生是否有必要采用不同的题型?结果表明,六种题型都能在一定程度上反映学生的口语水平,其中两种(“自我介绍”和“朗读”)对日韩和欧美学生有不同影响。这一结果为根据实际需要选择题型提供了较为可靠的标准。根据调查结果,我们设计了一套新的口语分班测试试卷,文章对新的试卷进行了具体说明,并且提出口语分班测试最好能包括一定的学能测试因素。

关键词 口语分班测试 题型 看图说话 朗读 自我介绍

口语分班测试一直是对外汉语教学界的一个薄弱环节,这种状况近年来有了一些改变。北京大学对外汉语教育学院的部分老师从交际策略、词汇量、篇章结构等方面对口语分班测试进行了探索,得到了一些有价值的结论(刘颂浩,2001)。本文在这些研究的基础上,重点探讨题型问题。文章分六部分:研究问题、研究方法、结果、研究结果的应用、讨论、结论。

一 研究问题

考试的题型取决于考试的要求、目标及可行性等。首先我们要明确口语分班测试的性质和目的。口语分班测试必然要考察学生现有的口头表达能力,这是它与口语水平测试的相同之处。但是,对分班测试来说,更重要的是为测试后的课堂教学服务。因此,必须要考虑分班后的课堂教学能否顺利进行。目前的口语教学,还不能脱离课本(至少北京大学是如此),这就涉及到课文认读的问题。即使某个学生的说话能力和班上其他同学一致,但如果汉字认读能力差距太大,也势必会影响到课堂教学效果。从这个角度讲,认读能力的重要性不容忽视。在相应的分班测试中,也需要兼顾认读能力。不过,认读能力属接受性技能,而说话能力属产出性技能,在口语测试中使用朗读,很可能对测试效度产生影响。我们在测试中使用朗读,也只是从教学实际情况出发作出的一种选择,并不意味着从理论上说应该或者一定要如此(参看下文四和五)。另外,分班测试还要考虑学生在所分到的班级里是否能

* 本文在第七届国际汉语教学讨论会(2002年8月,上海)上宣读。感谢匿名审稿人对本文初稿提出宝贵的修改意见,文中错误概由笔者负责。

最好地发挥自己的潜能,取得最好的学习效果。潜力大的学生分在稍高于他现有水平的班级里会进步更快。因此,了解考生的能力倾向也是口语分班测试的目的之一。

根据近年来参加和组织口语分班考试的经验,目前的分班测试存在着以下两个问题。

首先是题型的确定。据我们所知,口语测试题型一般都比较简单,高等 HSK 的口语考试采用“朗读”和“回答问题”两种题型,美国外语教学学会的口语能力测试(OPI)的题型是对话。相比较而言,北京大学对外汉语教育学院口语分班测试中采用过的题型则比较丰富,包括:自我介绍、回答问题、看图说话、话题表达、朗读短文或句子、对话等。那么,对于考察口语水平这个根本目的来说,这些不同的题型是否同样有效呢?如果需要使用组合题型,又应当如何进行选择呢?虽然从 1999 年以来,我们在题型的组合方面进行了不少尝试(刘颂浩,2001),但都只是凭感觉,并没有实证研究做基础。在对这个问题进行考察时,必须考虑分班测试的实际情况:规模较大,考生人数较多,考试时间有限(每个考生约六到七分钟)。理想的测试题型(或题型组合)应该能够在较短时间里,客观地反映出考生的实际口语水平,同时让考官比较轻松地对考生的口语水平作出准确评价。目前我们所使用的题型,还很难达到这一要求。

其次,对日韩和欧美学生是否要区别对待。众所周知,日韩学生的汉字认读能力较强,但口语表达能力和交际能力一般较差;而欧美学生则相反。前面提到,在口语分班测试中,需要兼顾认读能力。但是,如何在口语分班测试中更好地处理认读和表达这两种能力,仍然是一个棘手的问题。解决办法之一是针对不同背景的学生,使用不同的测试题型。不过,这只是一种理论上的推测,是不是真有必要,值得探讨。

因此,我们希望通过这次口语测试题型研究,回答下面两个问题:1)在口语分班测试中,什么样的题型或题型组合最有效?2)对不同背景的学生,是否有必要采用不同的题型?

二 研究方法

2.1 研究对象

参加这次调查的是北京大学对外汉语教育学院的 39 名留学生,时间是 2001 年 11 月底 12 月初。此时已经是下半学期,老师们对本班学生的口语水平已经有了清楚的了解,所以我们请老师们推荐班内中等水平的学生参加调查。39 名学生中,初级(3—5 班)16 名,中级(11—14 班)13 名,高级(18—20 班)10 名,其中日韩学生 21 人,欧美学生 18 人。

2.2 试卷构成

试卷共包括六种题型:自我介绍、对话、话题表达、看图说话、朗读和听读。其中前五种以前曾经采用过,听读为此次新增加的题型。每一题型的分值都是 10 分,全卷共 60 分。具体说明如下:

自我介绍。请考生简单介绍自己的情况,话题的范围不限。

考生与考官对话。这一部分借鉴了 OPI 的一些做法,如考官的问题根据考生的兴趣和经历而定,各个问题之间有一定的内在联系。同时,根据分班测试的要求和特点,在对话的内容和方法上做了相应的调整。

话题表达。我们列出了三个话题,考生可以任意选择一个。

看图说话。图画选自漫画集《父与子》,请考生先看一遍,看懂图画的意思后,再开始讲

述故事。具有一般生活常识的人都能看懂所选图画内容,即使对内容有不同的理解也没关系,评分时主要关注的是语言表达情况,而不是对内容的理解。

朗读。共有三个句子,分别用简体字和繁体字写出。考生先看一遍后开始朗读。三个句子根据长度和词语等级按从易到难的顺序排列,分值依次是3分、3分、4分。

听读。这是我们新增加的测试题型,以前的口语测试没有用过。不过,很多老师在口语课的教学中都使用过这一方法(李海燕,1998)。具体做法是,老师以正常语速说一个句子,学生听后再重复一遍。它主要考察听力、复述能力以及强记能力,我们觉得似乎可以从另一个角度体现考生的口语能力。调查时共听读三个句子:我的朋友每天都去打球;今天天气比昨天冷多了;要是不快点走的话,就来不及了。三个句子的分值依次是3分、3分、4分。

2.3 测试程序

调查采用一对一面谈的形式并录音,每个考生大约需要15—20分钟,录音材料共约12小时。考官是本文的三位作者,学生成绩在现场给出。两个星期后,再由另一位考官根据录音进行重评。

自我介绍、话题表达、看图说话、朗读、听读部分的评分标准见论文后的附录,对话部分参看刘颂浩(2001)。

三 结果

3.1 调查试卷的信度、效度分析

在对调查进行统计分析之前,有必要先考察一下本次调查所用试卷的信度和效度。口语测试属于主观性试题。“主观性试题一般比客观性试题有较高的效度。因为主观性试题易于根据考试需要直接命题”(桂诗春,1986:34)。但是因为评分过程中主观因素较大,评分标准不易做到完全一致,所以它的最大问题是信度不高。因此我们先对本次调查所用试卷进行信度、效度分析。这部分的分析包括三个方面:评估员信度、试题信度和区分效度。

先看评估员信度。计算评估员之间的信度的方法是两位评估员给同一批样本打分,然后计算两个分数的相关系数。我们计算了调查中两位评估员所评分数的相关系数,结果见表1。对样本的总体评估的相关系数为0.8653,这个结果对分班测试来说,是可以接受的。各分测验中相关系数最高、一致性最好的是听读,其他五个项目的一致性由高到低依次是:对话、话题表达、朗读、看图说话、自我介绍。由于同一位评估员对评分标准的把握也可能有前后不一致的情况,所以我们也随机抽取了8个样本进行了重评(两次评分隔了两个星期,重评时评分员不大可能记得以前打的分数),然后计算两次评分的相关系数,结果见表1。测试总分以及听读、看图说话、朗读、自我介绍的评估员内部信度比较高,这几部分两次评分的相关系数都达到0.9以上;但“自我介绍”的评分员间信度系数比较低,只有0.55。

表1 评估员间信度(N=39)和评估员内部信度(N=8)

	自我介绍	对话	话题表达	看图说话	朗读	听读	总分
两位评估员评分的相关系数	0.5478	0.8143	0.7628	0.6392	0.7431	0.9030	0.8653
一位评估员两次评分的相关系数	0.9129	0.8339	0.6708	0.9800	0.9197	0.9853	0.9789

再看试题的信度。一般用来计算信度系数的KR-20公式只适用于客观性试题,计算

主观性试题的信度系数需要使用 Cronbach 提出的 α 系数,方法是用一份试题中各大题分数的方差之和与整份试卷分数的方差进行计算(桂诗春,1986)。根据表 2 的数据,本次测试试题的 α 系数 = 0.9292,说明试题的信度很高,测试可靠。

表 2 39 名留学生的测试成绩

	自我介绍	对话	话题	看图	朗读	听读	总分
平均分	6.83	6.27	6.06	5.75	6.37	7.05	38.32
标准差	1.43	1.6	1.62	1.56	1.87	2	8.72
方差	2.0449	2.56	2.6244	2.4336	3.4969	4	76.0384

最后是试题的区分效度。本次调查的测试对象分别来自口语初、中、高三个不同层次的班级,这些人的口语水平是不同的,如果他们在各题型上的成绩有显著差异,那么就可以说明考试具有区分效度。表 3 的结果证明了这一点。

方差分析表明,初级、中级、高级三组的总分和各个分测验的成绩均存在显著差异。这说明调查所用六种题型都能有效地区分不同能力的考生。相对来说,三组在听读测试上的差异没有其他几项显著。原因可能是这次听读所用材料相对容易,导致该题的难易度值偏高,为 0.73。

表 3 初、中、高三组学生成绩的对比

		自我介绍	对话	话题	看图	朗读	听读	总分
平均分 (标准差)	初级 16 人	5.70 (1.26)	5 (1.26)	4.78 (1.25)	4.41 (1.12)	4.94 (1.44)	5.97 (1.9)	30.78 (6.49)
	中级 13 人	7.17 (0.96)	6.42 (0.77)	6.27 (1.09)	6.13 (0.89)	6.69 (1.39)	7.15 (1.54)	39.85 (3.67)
	高级 10 人	8.2 (0.54)	8.1 (0.88)	7.83 (0.69)	7.4 (0.83)	8.23 (1.09)	8.65 (1.67)	48.4 (3.94)
F 值		19.909	28.079	24.953	30.207	19.042	7.41	37.7
概值 (自由度 为 2 和 36)		<0.001	<0.001	<0.001	<0.001	<0.001	<0.005	<0.001
难易度		0.70	0.65	0.63	0.60	0.66	0.73	0.66

综上所述,本次调查所用试卷的信度和效度基本上达到了设计要求,可以认为是一份能反映学生口语水平的试卷,考官给学生评定的成绩也是可靠的,可以据此对考题做进一步的分析。

3.2 题型结构分析

对题型结构进行分析的目的是回答本文的第一个研究问题:什么样的题型或题型组合最有效?本文采用相关系数作为指标来考察这个问题。表 4 给出了本次调查所用题型的相关矩阵。

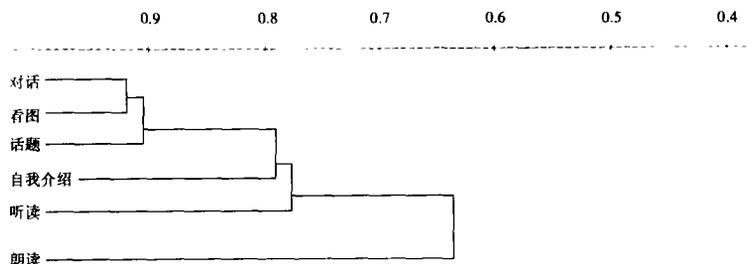
从表 4 可以看出,六个分测试与总分的相关系数都比较高(在 0.727 和 0.951 之间),说明这六个分测验都从不同的角度代表了全卷要考察的口语能力。其中与总分相关系数在 0.90 以上的依次是:看图说话、对话和话题表达。各分测验之间的相关系数比较高的是对话和看图说话(0.907)、话题表达和看图说话(0.905)、对话和话题表达(0.886)。朗读、听读与其它各题型基本上呈中等相关。

表4 测试各部分之间及各部分与总分的相关矩阵

	自我介绍	对话	话题表达	看图说话	朗读	听读	总分
自我介绍	1	0.796	0.776	0.787	0.566	0.595	0.852
对话		1	0.886	0.907	0.618	0.716	0.938
话题表达			1	0.905	0.527	0.743	0.923
看图说话				1	0.595	0.792	0.951
朗读					1	0.441	0.727
听读						1	0.831
总分							1

一般来说各题型之间的相关系数越高,则考试的内容结构就越合理。但是相关过高,说明各项考察的不是相对独立的语言能力因素,各个部分对于整个考试的效度增益不大。中等相关的题型则既有一致性又有个性,是从不同角度对语言能力进行考察(刘镰力,1997)。为了更好地观察这六种题型的关系,我们根据表4的结果用聚类谱系图将这六个测试项目进行了分类。聚类谱系图也叫树图(dendrogram),是利用距离或者相关系数的大小进行聚类的一种图分析方法。用这种方法分析的结果,与用基于主成分分析的因子负荷法进行分类的结果相近(李伟明,2001)。六个分测验的聚类情况见图1:

图1: 六个测试项目的谱系图



当取相关系数 $r=0.8$ 为分界线时,六项分测验可分为三类: {对话、看图、话题}, {自我介绍、听读}, {朗读}。第一类中的三个项目考察的都是学生口头综合表达能力,可以命名为综合表达类。还可以进一步从中选取一个,来代表这一类题型的典型特征。方法是计算相关系数的平方的均值,然后进行比较。结果是:对话 = 0.8041; 看图说话 = 0.8211; 话题表达 = 0.8028。因此,可以选择看图说话作为这一类的代表。第二类中只有两项(自我介绍和听读),可以任选一个作为代表。第三类当中只有朗读,说明这是独立的一类。上面三类题型分别代表的是综合表达能力、听力及复述能力(第二类中的听读)、朗读能力。据此,如果想使用单一题型,看图说话就是最佳选择;如果想使用两种题型,则可以有:看图说话 + 自我介绍,或者看图说话 + 听读;如需进一步增加题型,就要把朗读包括进来。

根据以上的分析,调查中选取的六种题型对测验口语能力都是有效的,这六种题型又可以分为三类: {对话、看图说话、话题表达}, {自我介绍、听读}, {朗读}。至于是否要组合,怎么组合,则取决于考试需要。

3.3 不同背景的学生在各题型测试中的表现

本文的第二个研究问题是,对不同背景的学生,是否有必要采用不同的题型? 我们通过

分析日韩和欧美两组学生在各个题型上的表现来回答这个问题。

参加口试的学生是经口语任课老师推荐的,他们在各自的班级里口语表达处于中等水平,所以每个级别的日韩学生和欧美学生的口语水平也基本一致。如果两组在某个分测试中成绩没有差别,那么可以认定这种题型可用;如果差异显著,就需要结合其他因素慎重考虑。在以前的口语分班测试中,有些老师觉得朗读部分对于欧美学生来说不公平,因为口头表达能力差的日韩学生可以在朗读部分得到比较高的分数,而口头表达能力比较强的欧美学生在朗读部分往往得不到分。那么实际情况如何呢?我们对日韩和欧美两组学生在六个分测验中的成绩进行了分析,结果见表 5。

表 5 日韩和欧美学生总成绩及分测试成绩的对比

		平均值	标准差	人数	总体标准差	t 值	单侧概值
自我介绍	日韩	6.45	1.33	21	1.38	-1.89	<0.05
	欧美	7.28	1.44	18			
对话	日韩	5.99	1.75	21	1.59	-1.196	>0.10
	欧美	6.60	1.39	18			
话题表达	日韩	5.75	1.78	21	1.60	-1.30	>0.10
	欧美	6.42	1.37	18			
看图说话	日韩	5.49	1.60	21	1.55	-1.145	>0.10
	欧美	6.06	1.50	18			
朗读	日韩	6.92	1.996	21	1.797	2.08	<0.025
	欧美	5.72	1.53	18			
听读	日韩	6.71	2.12	21	1.996	-1.139	>0.10
	欧美	7.44	1.84	18			
总分	日韩	37.35	9.64	21	8.77	-0.75	>0.10
	欧美	39.46	7.62	18			

表 5 显示,两组学生只在“自我介绍”和“朗读”这两部分中存在显著差异。欧美学生“自我介绍”的成绩比日韩学生好,原因大概与学生的性格有关。欧美学生比较外向,在自我介绍时愿意充分表现,因而得分较高。“朗读”部分日韩学生比欧美学生好,这和普通的经验是一致的。这个结果可以理解为,这两种题型凸现了两组学生口语水平的个别方面,在一定程度上夸大了他们整体水平的差异。

由于学生的汉字认读能力有一个逐渐发展的过程,因此我们也比较了口语水平对两组学生朗读的影响,结果如表 6:

表 6 不同程度的日韩和欧美学生在朗读部分的对比

	初级		中级		高级	
	日韩	欧美	日韩	欧美	日韩	欧美
平均分	5	4.875	7.464	5.792	8.833	7.313
标准差	1.69	1.246	0.488	1.6	0.97	0.375
人数	8	8	7	6	6	4

总体标准差	1.485	1.137	0.801
T 值	0.5978	2.6432	2.9421
自由度	14	11	8
单侧概值	>0.10	<0.025	<0.01

根据表6,在初级阶段日韩和欧美学生在朗读上没有明显差异,在中级、高级阶段两组学生的成绩分别在0.025和0.01的显著性水平上存在明显差异。由此可见,朗读部分两组学生的显著差异实际上只存在于中高级阶段。口语水平较低的学生,汉字的认读能力和口语整体能力基本一致;而中、高级阶段,汉字背景对学生的朗读成绩产生了影响,日韩学生比欧美学生具有优势。另一种可能是,欧美学生在中级阶段以后忽视了汉字学习,从而使认读能力滞后于口语整体能力。

根据上面的分析,可以得出结论:不同的题型确实对不同背景的学生有影响,自我介绍和朗读的影响最大。因此,如果考试目的是测量考生的整体口语水平,就不宜采用自我介绍和朗读这两种题型。

四 研究结果的应用

本次调查结果说明,试卷中所用的六种题型可以分为{对话、看图说话、话题表达}, {自我介绍、听读}, {朗读}三类;自我介绍和朗读会人为地拉大不同背景学生口语整体水平的差异。但是实际的口语分班测试,还不能完全依照这个结果。根据以上结论和实际需要,我们在2002年3月的入学口语分班测试中,设计了一套新的试卷。试卷共四个部分:自我介绍+简单问答(5分);看图说话(10分);朗读(10分);听读(7分)。全卷满分32分,测试时间约6-7分钟。具体讨论如下:

4.1 在调查中,“自我介绍”的评估员间信度较低,并且对不同背景的学生有不同的影响,所以口试试卷中可以不采用这一题型。但是在设计新的口试试卷时仍然保留这个项目并加上简单问答,目的是通过这个形式进行热身,缓解考生的紧张情绪,帮助考生更好地完成后面的测试,也使考官对考生的口语水平有一个初步的印象。为了尽量减少自我介绍这个题型对不同背景学生的影响,我们降低了它的分值。

4.2 参加分班考试的学生,从零起点到高等水平都有,差距很大,要通过6—7分钟的口试把这些学生的口语能力区分开来,试题的难度是否适中很重要。试题太难就区分不出水平较低的学生,过于容易则区分不出高水平的学生。我们在口语综合表达类的三个题型中选择了“看图说话”作为新试卷的第二个部分。这一题型给每个学生提供了一个展示口语水平的框架,避免了话题表达和对话中由于学生的阅读能力、听力能力差而造成的对口头表达的不利影响,同时也减轻了考官的负担。在本次调查中,看图说话与总分的相关也是最高的。

4.3 虽然在实验中得出了朗读对不同背景的学生有不同影响的结论,但学生的朗读能力对能否完成读课文等课堂教学活动至关重要。因此在新的试卷中仍然保留了朗读这一部分,但是对具体方法进行了调整。

以前朗读部分要读三个难度不同的句子,句子中包括不同难度等级的词语;评分标准也是综合性的,要考察的内容包括语音、语调、流利度、认字能力等。但实际测试中,这几个部

分互相影响,如果学生有不认识的字,在朗读时就必然影响到语音、语调和流利度。因此,我们对朗读这一部分的内容、形式和评分细则进行了调整。调整后的朗读部分把测量学生语音、语调、流利度的目的和考察认字能力的目的分离开来。学生先读一小段短文,短文中全部是甲级词,保证大部分学生都能认读,评分时只考察语音、语调和流利度;然后再读几个难度等级不同的词语,目的是考察学生的认字能力,评分时也十分简便,操作性较强,更接近客观性试题。两部分的分值各 5 分。经过这样的调整,测试目的和评分标准更明确了,同时,这种调整减少了汉字能力对其他能力的影响,对于欧美学生来说也更公平一些。

4.4 入学分班口试的目的是为口语课堂教学服务,因此应该尽可能地把具有同等学习能力的人分在一个班。前面谈到,口语分班测试的目的之一是了解考生是否具备在某个班级中学习的能力倾向。因此,分班测试在衡量学生现有口语水平的同时,最好也能关注潜在的学习能力。现代语言学能测试一般包括语音编码解码能力、语法敏感性、语言学习的归纳能力、语言的强记能力等(桂诗春,1986:16)。本次调查中采用的“听读”项目,能够考察学生的语音解码能力和强记能力,带有一定的学能测试的性质。本文第三部分的调查结果说明,“听读”在信度等各项指标上的表现都很突出,是一种大有可为的口试题型。因此,我们在新的口语试卷中加进了“听读”一项。

从实际操作来看,老师们对新的试卷反馈较好,认为操作简便,评分标准比较容易把握。从实际分班效果来看,也比较令人满意。

五 讨论:对朗读的反思

尽管“朗读”这一题型会对日韩和欧美学生产生不同的影响,但我们在分班测试中仍然采用了它,没有针对两类学生采用不同的题型。这主要是因为课堂教学的需要。北京大学对外汉语教育学院的口语课是混合编班,不同文化背景的学生在同一个班里学习,使用同样的教材。这就要求同一个班的学生有基本一致的朗读水平,否则就不能保证课堂教学的顺利进行。

本次调查发现,初级阶段日韩学生和欧美学生在朗读部分的成绩基本没有差别,中级、高级阶段的差异比较显著。为什么会出现这种现象,这种现象是否具有普遍性,需要进行专门研究。目前我们只能进行一些推测性的解释。我们的看法是,上述结果从某种程度上反映了学生对汉字认读的态度和教学当中的一些问题。在初级阶段,学生对基本汉字比较重视,并且有汉字课的专门学习,因此两类学生的认读差异不大。到了中高级阶段,教学中逐渐放松了对学生认读能力的要求,学生对汉字的需求也不如初级阶段那么迫切,日韩学生由于基础好,学起来容易,因而进步较快;而欧美学生则不愿意像基础阶段那样花大力气去学习汉字,进步也因此慢了下来。这一看法能否成立,还需要进行验证。

对上述有关朗读的调查结果的思考使我们认识到,在中、高级阶段,口语教学也仍然需要重视朗读训练。初、中级阶段的口头交际还局限在日常生活的范围,由于汉语的文白差异,不认识汉字似乎并不影响口头表达。但是到了高级阶段,口语能力就不能只限于日常生活范围。口头表达的词汇是否丰富、语法格式是否复杂多样、能不能在一些正式场合发表讲话等,是衡量口语水平的重要指标。这样,高水平口语表达中的词汇就会更多地向书面语靠近,两者的重合部分就会比较多。学生汉字的认读能力差,将直接影响到对这部分词汇的掌

握和使用,不利于学生在更高层次上提高口语水平。

虽然对朗读能力究竟代表什么,还没有一致的意见(刘颂浩,1999),但朗读能力能在一定程度上体现口语表达的综合水平,这也是毋庸置疑的。Urquhart & Weir(1998:26)认为,语言是一个抽象的系统,书面语言和口头语言是语言的两种平行的实现形式。在对外汉语教学领域,虽然不断有人反对“语文一体”的教学模式,强调要严格区分口语和书面语(比如吕必松,2003),但“语文一体”的教学模式却仍然广为流行,原因就在于现代汉语口语和书面语之间很难截然分开。陈建民(1984:79—80)指出,口语和书面语共同的句式占百分之八十五左右。我们赞成对外汉语教学中要区分口语和书面语,但同时也想指出,书面语的学习(主要是通过阅读,包括朗读来进行的)能够促进口语的发展,过分强调两者的区别不利于对外汉语教学的发展。

六 结论

本次调查对口试题型进行了专门的研究。依据调查结果,我们设计了新的口试试卷。通过调查和新一轮的实践,我们有以下几点认识。1)口语分班测试的性质与高等 HSK 口试、OPI 测试等其他口语水平测试不同,它应该包括一定的学能测试因素;2)题型组合可以根据需要采取多种形式。“自我介绍及简单问答+看图说话+朗读+听读”这样的题型结构对我们的分班口试来说是合理有效的。如果学生是单一背景的,当然可以灵活处理。比如,如果考生都是欧美学生或都是日韩学生,教学上也不需要,就可以不考朗读。本次调查中对题型类别的归纳,为合理安排题型提供了研究支持。3)朗读似乎很难作为口试的主要题型,但在口语教学中,即使到了中高级阶段,也仍然需要重视朗读训练。

在研究中我们也发现一些问题有待解决。比如,新的试卷中,朗读分成了两个部分,各有侧重。那么,对采用综合评分标准的看图说话,是否也要做同样的处理?因为学生的个体差异比较大,有的人较流利,有的人词汇语法好,另一些人则篇章意识强。高等 HSK 由于有充裕的评分时间,所以可以采用综合式与分解式相结合的办法(刘镰力,1997)。但分班测试显然没有这样的便利。这个问题还有待于进一步研究。

参考文献

- 陈建民(1984)《汉语口语》,北京:北京出版社。
- 陈田顺(1997)主观性试题的题型设计与阅读评分,见刘镰力(主编)1997。
- 桂诗春(1986)《标准化考试——理论、原则与方法》,南宁:广西高等教育出版社。
- 桂诗春(1997)《语言学方法论》,北京:外语教学与研究出版社。
- 李海燕(1998)谈初级汉语课文教学中的“听读”训练,《北大海外教育》第二辑,北京:北京大学出版社。
- 李伟明(2001)《多元描述统计方法》,上海:华东师范大学出版社。
- 刘镰力(1995) HSK(高等)的内容效度与题型开拓,《汉语水平考试研究论文选》,北京:现代出版社。
- 刘镰力主编(1997)《汉语水平测试研究》,北京:北京语言文化大学出版社。
- 刘颂浩(1999)对9名日本学生的误读分析,《语言教学与研究》第2期。
- 刘颂浩(2001)口语测试的组织与实施探索,《华文教学与研究》第3期。
- 吕必松(2003)汉语教学路子研究刍议,《暨南大学华文学院学报》第1期。
- 文秋芳(1999)《英语口语测试与教学》,上海:上海外语教育出版社。

Urquhart, A. H. & Weir. C. J. (1998) *Reading in a second language: process, product, and practice*. London: Longman.

附录:调查所用五种题型的评分标准

说明:以下的评分标准基本以 1 分为单位,但根据情况,可以使用 0.5 分。

一 自我介绍

级别	分数	篇章结构	话题
初级	1	1	姓名、国籍、年龄、家庭、爱好、专业、旅行、业余生活等
		2	
	2	3	
		4	
中级	3	5	姓名、爱好、学汉语的原因、工作、专业等
		6	
	4	7	
		8	
高级	9	姓名的来历、经历等	
	10		

二 话题表达

级别	分数	篇章结构	内容	功能
初级	1	1	只能说几句,表达的意思十分模糊	
		2		
	2	3		
		4		
中级	3	5	意思比较清楚,但很不深入	叙述、比较
		6		
	4	7	说明的内容比较完整清楚,在某些地方较深入	叙述、比较描述
		8		
高级	9	篇章层次清楚,语言方面的错误比较少	讨论具体深入,叙述说明完整清楚	能进行抽象的议论
	10			

三 看图说话

级别	分数	篇章结构	内容
初级	1	1	故事很不完整,没有看过该图的人听不明白
		2	
	2	3	
		4	
中级	3	5	能描述图画的基本内容,细节较少,使没看过图画的人基本了解故事经过
		6	
	4	7	
		8	
高级	9	对图画内容描述更加细致,因果关系较清楚,能就故事发表抽象的议论	
	10		

四 朗读句子

句子	得分
A:我姐姐对滑冰一点儿也不感兴趣。	0.5 1 2 3
B:屋子里铺上红地毯以后,显得既温暖又舒适。	0.5 1 2 3
C:他擅长拉小提琴,曾多次在国际大型比赛中取得过优异的成绩。	0.5 1 2 3 4

评分标准:

(1)只认识最常见的字,得 0.5 分。(2)认识大部分字,朗读以字为单位,得 1 分。(3)认识所有的字,以词为单位朗读,但停顿不明确,得 2 分(C 句得 3 分)。(4)认识所有的字,以词为单位朗读,停顿、重音和语气都正确,得 3 分(C 句得 4 分)。

五 听读句子

句子	得分			
A:我的朋友每天都去打网球。	0.5	1	2	3
B:今天天气比昨天冷多了。	0.5	1	2	3
C:要是不快点儿走的话,就来不及了。	0.5	1	2	3 4

评分标准:

(1)只能说出最前边或者最后边的几个字,得0.5分。(2)能说出大部分词,意思和原句有些出入,得1分。(3)能说出绝大部分的词,意思也基本符合原文,但是丢失了一些虚词,得2分(C句得3分)。(4)能说出所有的词,停顿、重音和语气都正确,得3分(C句得4分)。

作者简介

李海燕,女,北京大学对外汉语教育学院讲师。

蔡云凌,女,北京大学对外汉语教育学院讲师。

刘颂浩,男,北京大学对外汉语教育学院副教授。

北京地区第三届对外汉语教学学术讨论会召开

2003年10月18日,北京地区第三届对外汉语教学学术讨论会在中国人民大学召开。会议由中国人民大学对外语言文化学院承办。国家对外汉语教学领导小组办公室副主任姜明宝、中国人民大学副校长林岗、北京语言大学副校长霍明杰、世界汉语学会会长陆俭明、对外汉语教学界著名专家学者胡明扬、吕必松以及来自北京地区22所大学及相关单位的同行、研究生近400人参加了大会。

大会开幕式由北京语言大学杨惠元主持,北京大学李晓琪代表会议筹备组致开幕辞。中国人民大学副校长林岗代表承办单位对大会的召开表示热烈的祝贺。国家汉办副主任姜明宝在致辞中指出,在对外汉语教学事业蓬勃发展的大好形势下,国家汉办将把教材建设、网络教学和人才培养作为工作的重点,并即将启动“汉语桥工程”,力争把对外汉语教学事业推向一个新的高度。陆俭明在致辞中提出,新世纪对外汉语教学的大发展,北京要起到龙头作用,首先要在学科建设上起带头作用。胡明扬、吕必松也向大会致辞,并分别强调要继承我国语言教学的优良传统和加强对外汉语教学的实验研究。对外汉语教学界知名专家赵金铭、鲁健骥、刘珣、郭振华等应邀出席了开幕式,并参加了分组讨论。

本次大会共收到近30所学校对外汉语教师的论文250余篇。北京语言大学崔永华教授、中国人民大学李泉教授、北京大学杨德峰副教授、北京师范大学王魁京教授分别以《略论对外汉语教学的教学设计》、《论对外汉语教材的针对性》、《试论“没有一个人不知道”及其相关句式》、《从语言输出的角度看海外华人学生心理词典的特点》为题做了大会发言。在分组讨论中,与会专家学者就汉语语法、语音、词汇和汉字研究、汉外对比研究、汉语认知和测试研究、教学管理研究、教学法研究、课程研究、教材研究、网络 and 多媒体教学研究、汉语学习和习得研究等进行了讨论。

(杜 健供稿)

tures of compound words, such as words formed with morphemes, structures and semantic relevance, ect. and 2) Foreign learners may employ two different ways of acquisition of compound words, namely, compounding from morphemes and recognising compound words as a whole. We hold that the former should be in the dominant.

Key words compound words error compound words target words the Chinese Inter-language Corpus System

A study of the item types of oral placement tests

..... **LI Haiyan, CAI Yunling and LIU Songhao, p79**

This paper has made an analysis of the six item types in previous oral placement tests. It aims to answer two questions: 1) Which of the six types are more effective than others? and 2) Is it necessary to use different item types for students of different backgrounds? The result of the analysis shows that all the six types can reflect, to some extent, the speaking proficiency of the students and two of them (self-introduction and reading aloud) have effects on Japanese and Korean students that are different from those on European and American students. This result has provided us with reliable criteria for selecting different item types for different students. We have designed a set of oral placement tests, and it is our view that some aptitude test items should be included. The paper has made an analysis of reading aloud test and emphasised that reading aloud must be attached importance in the oral work at the intermediate and advanced levels, however, it is not appropriate to be used as an examination item type, at least not as a major item type.

Key words oral placement test item types speak on a picture reading aloud self-introduction dialogue

A statistical analysis of citations in *Chinese Teaching in the World*

between 1987 ~ 2001 **SONG Ruoyun, p107**

This paper is a statistical analysis of 5,952 citations in *Chinese Teaching in the World* (1987 - 2001). With the counting, comparing, summarising and assessment of the citations, it aims at finding their quantitative features and implied values. It provides quantitative norms on the developmental trends, major academic writings and specialists, structure of knowledge and the evaluation of performances of researchers. The results reflect the status quo and tendency of the teaching and research of Chinese as a foreign language, as CTW is one of the major journals in the field.

Key words *Chinese Teaching in the World* citation quantitative norm